

RESEARCH ARTICLE

Open Access

Identification of a novel Parkinson's disease locus via stratified genome-wide association study

Erin M Hill-Burns¹, William T Wissemann¹, Taye H Hamza¹, Stewart A Factor², Cyrus P Zabetian³ and Haydeh Payami^{1,4*}

Abstract

Background: Parkinson's disease (PD) is complex and heterogeneous. The numerous susceptibility loci that have been identified reaffirm the complexity of PD but do not fully explain it; e.g., it is not known if any given PD susceptibility gene is associated with all PD or a disease subtype. We also suspect that important disease genes may have escaped detection because of this heterogeneity. We used presence/absence of family history to subdivide the cases and performed genome-wide association studies (GWAS) in Sporadic-PD and Familial-PD separately. The aim was to uncover new genes and gain insight into the genetic architecture of PD.

Results: Employing GWAS on the NeuroGenetics Research Consortium (NGRC) dataset stratified by family history (1565 Sporadic-PD, 435 Familial-PD, 1986 controls), we identified a novel locus on chromosome 1p21 in Sporadic-PD ($P_{\text{NGRC}} = 4 \times 10^{-8}$) and replicated the finding ($P_{\text{Replication}} = 6 \times 10^{-3}$; $P_{\text{Pooled}} = 4 \times 10^{-10}$) in 1528 Sporadic-PD and 796 controls from the National Institutes of Neurologic Disease and Stroke (NINDS) Repository. This is the fifth PD locus to be mapped to the short arm of chromosome 1. It is flanked by *S1PR1* and *OLFM3* genes, and is 200 kb from a multiple sclerosis susceptibility gene. The second aim of the study was to extend the stratified GWAS to the well-established PD genes. *SNCA*_rs356220 was associated with both Sporadic-PD ($\text{OR} = 1.37$, $P = 1 \times 10^{-9}$) and Familial-PD ($\text{OR} = 1.40$, $P = 2 \times 10^{-5}$). *HLA*_rs3129882 was more strongly associated with Sporadic-PD ($\text{OR} = 1.38$, $P = 5 \times 10^{-10}$) than Familial-PD ($\text{OR} = 1.12$, $P = 0.15$). In the *MAPT* region, virtually every single nucleotide polymorphism (SNP) had a stronger effect-size and lower P-value in Familial-PD (peak $P = 8 \times 10^{-7}$) than in Sporadic-PD (peak $P = 2 \times 10^{-5}$).

Conclusions: We discovered and replicated a new locus for Sporadic-PD which had escaped detection in un-stratified GWAS. This demonstrates that by stratifying on a key variable the power gained due to diminished heterogeneity can sometimes outweigh the power lost to reduced sample size. We also detected distinct patterns of disease associations for previously established PD susceptibility genes, which gives an insight to the genetic architecture of the disease and could aid in the selection of appropriate study population for future studies.

Keywords: GWAS, Parkinson's disease, *SNCA*, *MAPT*, *HLA*, Genetic heterogeneity, Secondary GWAS, Stratified GWAS, Chromosome 1p

Background

PD is heterogeneous. Despite the great strides made recently, we still do not have a clear picture of the genetic architecture of PD, partly because not all the genes have been identified, and partly because we do not know if a given gene is associated with all or a subtype of PD. An

important outcome of GWAS will be to use the vast information content that has been gained [1-7] to define disease subtypes based on their genetic associations. As a starting point, we posit that using surrogates for the underlying heterogeneity may help define the disease subtypes that each gene is associated with and may also reveal genes that were previously masked by this heterogeneity. Here, we use family history as a surrogate. The majority of PD (70%-85%) is non-familial (henceforth Sporadic-PD). The remaining 15%-30% of PD patients have a positive family history (Familial-PD), but rarely do their kindreds display a Mendelian inheritance pattern

* Correspondence: hpayami@wadsworth.org

¹Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, NY, USA

⁴Department of Biomedical Science, School of Public Health, State University of New York, Albany, NY, USA

Full list of author information is available at the end of the article

(Mendelian-PD). Mendelian-PD has been linked to pathogenic mutations in *SNCA*, *LRRK2*, *PARK2*, *PINK1*, *DJ1*, *ATP13A2* and *VPS35* [8-16]. The vast majority of Familial-PD remains idiopathic. The genetic distinction between idiopathic Familial-PD and Sporadic-PD, if any exists, is unknown.

Idiopathic PD involves complex interactions between the genome and environmental exposures [6,7,17,18]. It is operationally assumed that the same set of susceptibility genes predispose to Familial and Sporadic-PD. In fact, GWAS have successfully uncovered numerous susceptibility loci without separating the subtypes [1-7]. We hypothesized that Familial-PD and Sporadic-PD have different genetic structures. We acknowledge that since genetic disease can manifest without a family history due to incomplete penetrance (e.g., *LRRK2* mutations [19]), and environmentally-induced disease can cluster in families due to common exposure, there must exist an invisible overlap between Sporadic and Familial-PD. However, they might differ in the relative burden of incompletely penetrant Mendelian genes vs. genes that confer susceptibility to environmental causes. There is evidence in the literature that supports this notion: consider three well-established PD-associated genes: *GBA*, *LRRK2*, and *HLA*. *GBA* mutations are significantly more common in Familial than in Sporadic-PD [20]. *LRRK2* G2019S is also significantly more common in Familial than in Sporadic-PD [21]. *HLA*, on the other hand, is more strongly associated with Sporadic-PD than with Familial-PD [3]. There are often no overt phenotypic differences between these genetic subtypes. Subtle clinical differences were only detected after the genes were identified and subtypes were defined genetically. This in itself underscores the importance of finding the genes.

Here we report the first GWAS stratified by Sporadic and Familial-PD which identified a previously unknown PD susceptibility gene in Sporadic-PD. We also present evidence for distinct patterns of associations for several well-established PD susceptibility loci with familial and sporadic subtypes.

Methods

This study was approved by institutional review boards at the participating institutions: Albany Medical Center, Emory University, Kaiser Permanente Northwest Division, New York State Department of Health, Oregon Health & Sciences University (OHSU) and the Department of Veterans Affairs VA Puget Sound Health Care System (VAPSHCS). All study participants gave informed consent. All patients and most control subjects gave written signed consent; a portion of control subjects, recruited at OHSU and VAPSHCS, who wished to remain anonymous read the written informed consent and gave verbal consent as approved by the institutional review

boards at OHSU and VAPSHCS respectively. All participants were adults and gave consent on their own behalf. No parents or guardians were asked to consent for the subjects. Subjects were from NGRC and included 2000 persons with PD (435 Familial-PD, 1565 Sporadic-PD) and 1986 controls (Additional file 1). PD was diagnosed by movement disorder specialists using UK/NINDS diagnostic criteria [22]. Controls were free of neurodegenerative disease, 340 of them were examined by a neurologist. All 3986 subjects were confirmed as genetically unrelated ($PI_HAT \leq 0.15$). Familial-PD was designated for cases with one or more first or second-degree relatives with PD; Sporadic-PD was all other cases. There was no significant difference between Familial and Sporadic cases in age at recruitment, age at onset, gender, percentage of Ashkenazi Jews, and the inverse association of smoking and coffee with PD (Additional file 1).

The genome-wide genotype data were generated by our group and are publicly available on dbGaP (<http://www.ncbi.nlm.nih.gov/gap>). Standardized subject selection criteria, protocols and subject characteristics were used which are published [3] and are also available online (<http://www.ncbi.nlm.nih.gov/gap>) with complete genotype and phenotype data on the entire NGRC cohort (phs000196.v2.p1). Genome-wide genotyping was performed on DNA from whole blood using the Illumina HumanOmni1-Quad_v1-0_B array [3]. 811,597 SNPs passed quality-control (minor-allele frequency ≥ 0.01 , call-rate $\geq 99\%$, Hardy-Weinberg $P \geq 10^{-6}$, allele-frequency difference in men vs. women ≤ 0.15 , missing rate in cases vs. controls $P \geq 10^{-5}$). An additional 6.4 million SNPs with frequency ≥ 0.01 were imputed with high fidelity (info-score $\geq 90\%$) using the 1000 Genomes Phase I integrated variant set release v3 and the IMPUTE v.2.2.2 software [23]. The analyses were conducted on the total of 7.2 million SNPs. For rs2338971 (the top signal on chromosome 1), the call rate for individuals with imputed genotype probability $\geq 90\%$ was 0.98 in cases and 0.98 in controls. Genotype frequencies in controls for rs2338971 were in Hardy Weinberg equilibrium ($P = 0.24$).

For replication, we used cases and controls from the NINDS Human Genetics DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>). We obtained the NINDS GWAS data [2] from dbGaP (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000089.v3.p2). We used only white non-Hispanic subjects. The sample size was 924 cases (621 Sporadic-PD, 303 Familial-PD) and 797 controls. We imputed the two SNPs of interest on chromosomes 1 and 8 with high confidence (info-score ≥ 0.97). In addition, we had purchased 1490 PD DNA samples (1025 Sporadic, 465 Familial) from the NINDS Repository. We directly genotyped the DNA samples using TaqMan. The GWAS dataset and our DNA samples had only 120 cases in common. This allowed us

to combine the non-overlapping segments of the two datasets to attain a larger sample size, as well as an opportunity to validate imputation vs. genotyping results using the overlapping 120 cases. The genotypes, TaqMan vs. imputation, for the top signal at chromosome 1 matched 100% for these 120 individuals. Call rates were 0.95 for genotyped cases, 0.97 for imputed cases with genotype probability $\geq 90\%$, and 0.98 for imputed controls with genotype probability $\geq 90\%$. Controls were in Hardy-Weinberg proportions ($P = 0.11$). The final sample size for replication was 2235 cases (1528 Sporadic, 707 Familial) and 796 controls, all from the NINDS Repository.

Statistical analyses for GWAS were performed using ProbABEL v.0.1-9d [24] adjusting for age at blood draw, sex and two principal components; and using R version 3.0.1 (<http://www.r-project.org/>) for replication adjusting for sex and study. Differences in allele frequencies of Sporadic vs. Familial cases were tested using logistic regression. Conditional analysis was conducted using logistic regression. P values shown for replication of chromosome 1 signal are one-sided, due to the directionality of the hypotheses being tested [25]. Linkage disequilibrium (LD) was assessed using Haploview V-4.2 [26] and LocusZoom [27].

We checked the 44 SNPs in the chromosome 1 peak (those with $P < 10^{-5}$) for evidence of association with gene expression in expression quantitative trait loci (eQTL) databases (<http://www.sanger.ac.uk/resources/software/genevar/>, <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>, <http://gbrowse.csbio.unc.edu/cgi-bin/gb2/gbrowse/seeqtl/>, <http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi>, <http://www.scandb.org/newinterface/about.html>) and in published eQTL studies for variants associating with gene expression in the brain [28-32].

Results

Known loci

When GWAS data were analyzed without stratification (All-PD with 2000 cases, 1986 controls), we recovered *SNCA* as the strongest signal; *HLA*, which also reached genome-wide significance; and *MAPT*, which had a clear peak but fell below the significance threshold (Figure 1A, Table 1). It is well established that polymorphisms in *SNCA*, *MAPT*, and *HLA* are associated with PD; it is not known however if their effects are ubiquitous across all PD or stronger in Sporadic or Familial subtype.

The top hit for *SNCA* region was rs356220 at the 3' of the gene (All-PD: OR = 1.38, $P = 3 \times 10^{-11}$). This variant was also the most significant marker in both Sporadic-PD (OR = 1.37, $P = 1 \times 10^{-9}$) and Familial-PD (OR = 1.40, $P = 2 \times 10^{-5}$). There was no difference in the *SNCA* rs356220 allele frequencies in Familial and Sporadic-PD ($P = 0.85$).

The top hit for *HLA* was rs3129882 in intron 1 of *HLA-DRA* (All-PD: OR = 1.31, $P = 3 \times 10^{-8}$). This association was

strong in Sporadic-PD (OR = 1.38, $P = 5 \times 10^{-10}$) but weak and statistically non-significant in Familial-PD (OR = 1.12, $P = 0.15$). The difference between Sporadic and Familial-PD in the frequency of *HLA-DRA* rs3129882 alleles was significant ($P = 6 \times 10^{-3}$).

Within the *MAPT* region, defined from 43.5 Mb to 44.9 Mb on chromosome 17 (Human Genome Build 37) and including *PLEKHM1*, *MAPT*, *NSF* and *WNT3* genes, 2,365 SNPs gave evidence for association with PD ($P < 10^{-5}$ in All-PD). The effect sizes for the 2,365 SNPs were always greater in Familial-PD ($0.56 \leq \text{OR} \leq 0.68$) than for Sporadic-PD ($0.76 \leq \text{OR} \leq 0.81$). Moreover, nearly every SNP (2,363 of 2,365) achieved higher statistical significance in Familial-PD ($2 \times 10^{-4} \geq P \geq 8 \times 10^{-7}$) than in Sporadic-PD ($8 \times 10^{-4} \geq P \geq 2 \times 10^{-5}$). The GWAS signal for the *MAPT* region peaked at $P = 8 \times 10^{-7}$ in Familial-PD vs. $P = 2 \times 10^{-5}$ in Sporadic-PD. The sample size for Sporadic-PD was 3-times larger than Familial-PD and therefore power was not a limiting factor in this case because the association was strong in Familial-PD despite the smaller sample size.

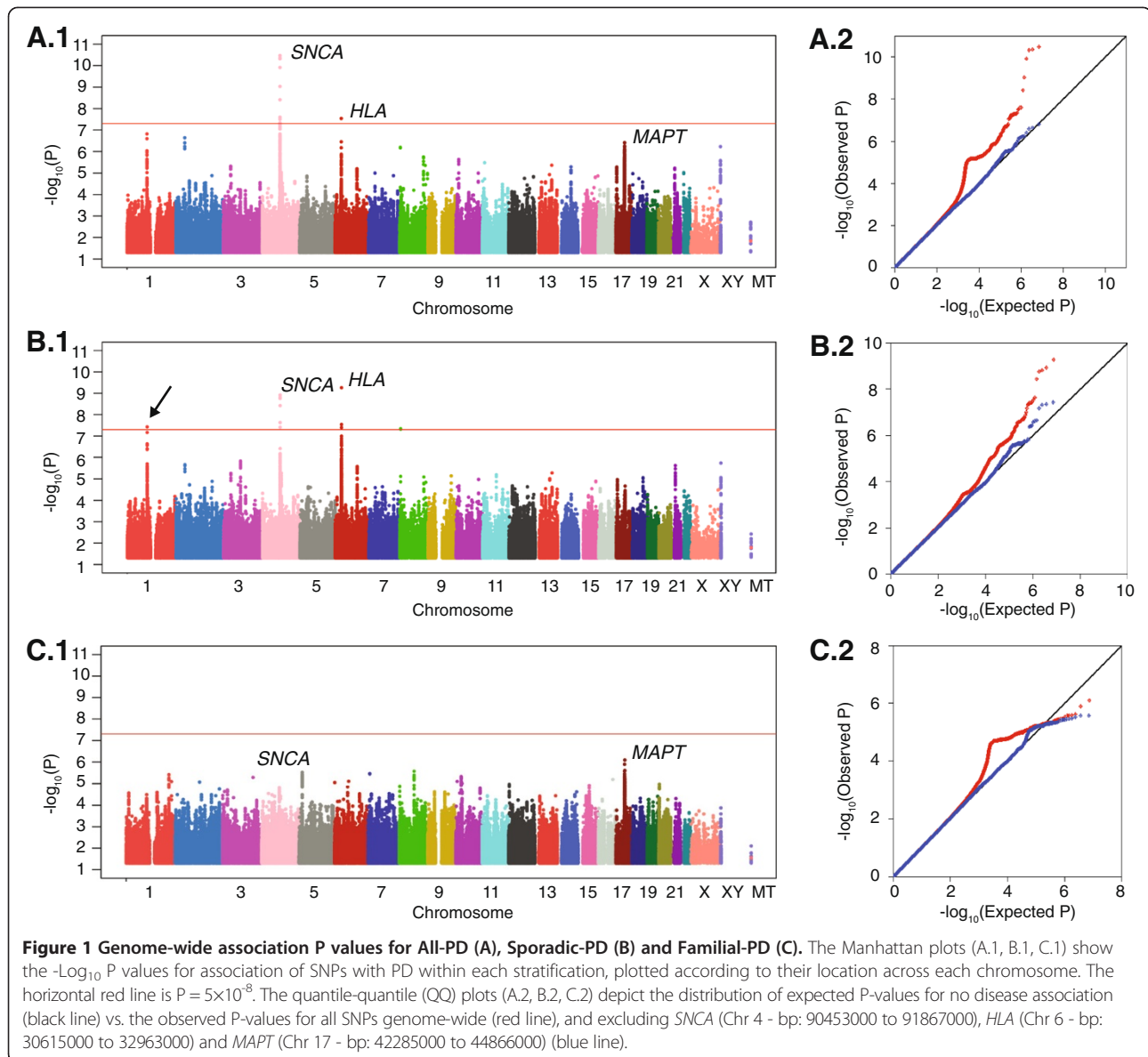
New locus

The GWAS on Sporadic-PD (1565 cases vs. 1986 controls) revealed two signals, (Figure 1B, Table 1), one on the short arm of chromosome 1 (a SNP-dense peak with top signal at rs2338971, $P = 4 \times 10^{-8}$) and another on chromosome 8 (a single SNP, rs12681349, $P = 5 \times 10^{-8}$). The GWAS on Familial-PD (435 cases vs. 1986 controls) did not reveal any statistically significant signals (Figure 1C).

We tested the new Sporadic-PD signals in an additional 1528 Sporadic-PD cases, 707 Familial-PD cases and 796 controls from the NINDS Repository (Table 1). The signal on chromosome 1 replicated. In Sporadic-PD, the top chromosome 1 SNP, rs2338971, gave $P_{\text{Replication}} = 6 \times 10^{-3}$ with OR = 0.81 which was in the same direction as NGRC and therefore boosted the significance of the combined data to $P_{\text{Pooled}} = 4 \times 10^{-10}$. The chromosome 8 signal did not replicate (Table 1).

The confirmed signal maps to the short arm of chromosome 1 at p21. The signal is a strong peak with 44 SNPs that achieved $4 \times 10^{-8} \leq P < 10^{-5}$ for association with Sporadic-PD in the NGRC dataset. We defined the new locus based on the location of these 44 SNPs as a 141 kilo base-pair (kb) region from base-pair (bp) 101,872,292 to 102,013,715.

We performed sensitivity analysis testing association of the top SNP, rs2338971, within subgroups of Sporadic-PD classified by sex, age at onset, genotype (*SNCA*, *HLA* and *MAPT*), exposure (smoking, caffeinated-coffee, non-steroidal anti-inflammatory drugs use), recruitment site, European countries of ancestral origin, and Ashkenazi Jewish heritage. Association of rs2338971 with Sporadic-PD was robust in every stratum with no evidence of heterogeneity across strata (Additional file 2).



Four PD loci have been mapped to chromosome 1p, namely, *PARK6/PINK1*, *PARK7/DJ1*, *PARK9/ATP13A2*, and *PARK10* (Figure 2A). The new locus is more than 45 mega base-pairs (Mb) away from the closest known PD locus, which is *PARK10*. We tested LD between the SNPs that reached $P < 10^{-5}$ for association with Sporadic-PD (44 SNPs) and SNPs that were in or within ± 10 kb of *DJ1* (66 SNPs), *PINK1* (158 SNPs), *ATP13A2* (107 SNPs) and *PARK10* (15,796 SNPs). There was no correlation between the new locus and the known loci ($r^2 = 0$). We therefore conclude that the signal identified here represents a previously unknown PD susceptibility locus.

The associated region maps onto a poorly-annotated gene, *huseyo* (Figure 2B). *Huseyo* is transcribed [33,34] but little is known about the putative protein. rs10493953,

which gave $OR = 0.64$ with $P = 1 \times 10^{-5}$ for association with Sporadic-PD is predicted to cause a non-synonymous substitution (A/G \rightarrow Q13R) in the putative protein. However, when we conditioned on the strongest signal (rs2338971), the signal for the amino-acid changing SNP was lost ($P = 10^{-5}$ dropped to $P = 0.07$). The top SNP, rs2338971, also dropped in significance but was not abolished when conditioned on the amino-acid changing SNP ($P = 4 \times 10^{-8}$ to $P = 2 \times 10^{-4}$). These results suggest neither of these SNPs can explain the association signal entirely, rather they are probably tagging an as yet unknown disease-associated factor in the region.

There are several genes in the region (Figure 2B) including *VCAM1*, *SLC30A7*, *SIPR1* and *OLEFM3* that could be considered plausible candidate genes for PD based on

Table 1 Results of GWAS conducted in All-PD and stratified by Sporadic-PD and Familial-PD

CHR Gene	SNP	Base pair position*	Minor/major allele	Allele frequency				Control	Association with PD								
				Case					All-PD			Sporadic-PD			Familial-PD		
				All-PD	Sporadic	Familial			OR	SE	P	OR	SE	P	OR	SE	P
Known highly significant loci																	
4 <i>SNCA</i>	rs356220	90641340	T/C	0.440	0.439	0.443	0.364	1.38	0.07	3×10 ⁻¹¹	1.37	0.07	1×10 ⁻⁹	1.40	0.11	2×10 ⁻⁵	
6 <i>HLA</i>	rs3129882	32409530	G/A	0.459	0.470	0.418	0.395	1.31	0.06	3×10 ⁻⁸	1.38	0.07	5×10 ⁻¹⁰	1.12	0.09	0.15	
17 <i>MAPT</i>	rs199498	44865603	C/T	0.190	0.197	0.162	0.236	0.74	0.05	2×10 ⁻⁶	0.78	0.05	2×10 ⁻⁴	0.59	0.06	8×10 ⁻⁷	
New Signals in NGRC																	
1	rs2338971	101880005	T/C	0.187	0.179	0.213	0.232	0.74	0.04	2×10 ⁻⁷	0.71	0.04	4×10 ⁻⁸	0.86	0.08	0.12	
8	rs12681349	4277990	T/C	0.376	0.369	0.403	0.426	0.78	0.04	7×10 ⁻⁷	0.75	0.04	5×10 ⁻⁸	0.88	0.07	0.10	
Replication																	
1	rs2338971	101880005	T/C	0.184	0.182	0.188	0.212	0.82	0.06	5×10 ⁻³	0.81	0.07	6×10 ⁻³	0.83	0.08	0.03	
8	rs12681349	4277990	T/C	0.427	0.420	0.442	0.419	1.02	0.06	0.38	0.99	0.06	0.41	1.09	0.08	0.13	
Pooled NGRC and Replication																	
1	rs2338971	101880005	T/C	0.185	0.181	0.197	0.227	0.76	0.03	5×10 ⁻¹⁰	0.74	0.04	4×10 ⁻¹⁰	0.82	0.05	2×10 ⁻³	

*Genome Build 37.

their known functions (see Discussion). To explore the possibility that the new signal is tagging a linked gene, we examined the pattern of LD in the region (Figure 2C). Using rs2338971 as the anchor we found that strong LD ($r^2 \geq 0.8$) extended only 8 kb telomeric and 16 kb centromeric; moderate LD ($r^2 \geq 0.5$) extended 8 kb and 100 kb and marginal LD ($r^2 \geq 0.2$) extended 30 kb and 157 kb. Therefore, even the closest genes to this signal (*SIPRI* at ~165 kb telomeric and *OLFM3* at ~250 kb centromeric) were outside the region of marginal LD with it. To be specific, there was no LD between the 44 SNPs of the new signal and SNPs in or within ± 10 kb of *OLFM3* (86 SNPs tested, $r^2 = 0$) or *SIPRI* (11 SNPs tested, $0 \leq r^2 \leq 0.08$).

The block of SNPs that showed the strongest association with Sporadic-PD ($5 \times 10^{-8} < P < 5 \times 10^{-7}$) overlaps with a DNA sequence that contains strongly marked regulatory regions in the ENCODE database [34] (Figure 2D). We searched five eQTL databases and published expression data on human brain (see Methods) and did not find evidence for any of the SNPs being an eQTL.

Discussion

Although PD is recognized as a complex and heterogeneous disease, it has been treated as a single entity in most prior GWAS. While GWAS have been enormously successful, the discoveries to date are only a fraction of the information content that these large and expensive datasets encompass. In this paper, we show that accounting for aspects of disease heterogeneity, in this case family history, could give new insights into the genetic architecture of disease. Stratifying by family history, we discovered and replicated a new locus for Sporadic-PD which had escaped detection in un-stratified GWAS. We also detected distinct

patterns of disease associations for some of the previously established PD susceptibility genes. We did not detect any new signals for Familial-PD, which is not surprising because the sample size was small. In order for this approach to work, the power gained by reducing heterogeneity must outweigh the power lost to decreased sample size.

This is the fifth PD locus to be mapped to chromosome 1p, which including *DJ1*, *ATP13A2*, *PINK1*, and *PARK10*, was already coined as a PD hot spot. A multiple sclerosis (MS) susceptibility locus [35] also maps to chromosome 1p, only ~200 kb away from the signal for PD. The gene for MS has not been identified. There are several plausible candidate genes for PD in the region (Figure 2). *SIPRI* (sphingosine-1-phosphate receptor 1) encodes a lipid G protein-coupled receptor involved in cell-cell adhesion, which affects differentiation of endothelial cells and has a crucial role in immune response [36,37]. *VCAM1* (vascular cell adhesion molecule 1) mediates leukocyte-endothelial cell adhesion and signal transduction, and is critical for adult neurogenesis by maintaining the structure and function of the adult forebrain subventricular zone where neuronal stem cells give rise to neurons [38]. *SLC30A7* (solute carrier family 30 member 7) encodes the zinc transporter 7 protein (ZNT7) [39]. Zinc is required for synaptic neurotransmission and can also act as an antioxidant [40]. Zinc accumulates selectively in the substantia nigra of PD brains [41,42], and in rodents, is shown to both enhance and reduce excitability of dopaminergic neurons [43]. ENCODE data [34] show strong markings within the most significant region of PD association that are predicted to be transcription factor binding sites, DNase hypersensitivity clusters, and H3K27ac histone marks, which indicate

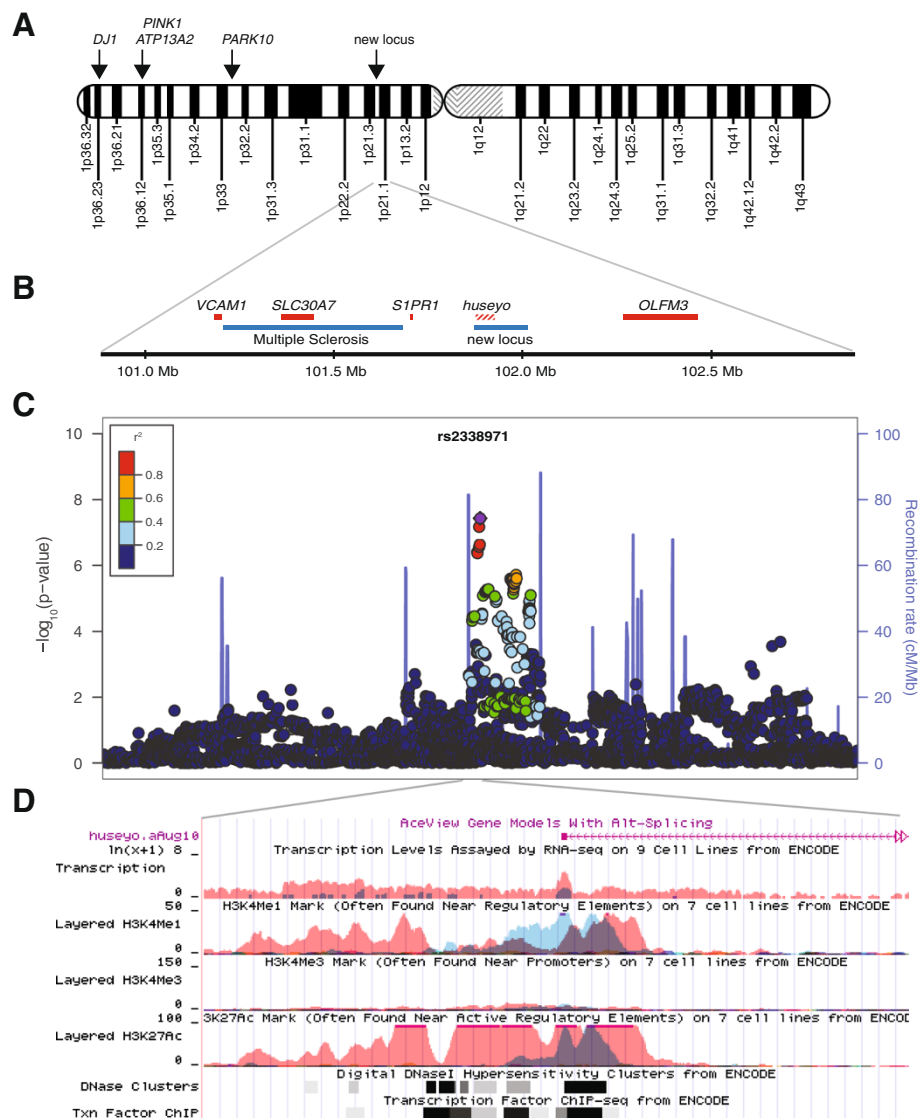


Figure 2 New region of association with Sporadic-PD. (A) Locations of PD-associated loci on the short arm of chromosome 1. **(B)** Enlargement of chromosomal region centered on the new signal and extending 1 Mb in each direction. Red bars are known genes. Hashed red bar, *huseyo*, is poorly annotated. Not all known/putative genes in the region are displayed. Blue bars are disease loci that have been mapped but the genes have not been characterized yet. **(C)** The LD structure of Sporadic-PD associated SNPs in the 2 Mb region. SNPs within 1 Mb on either side of top SNP, rs2338971, are plotted, showing $-\log_{10} P$ values for their association with Sporadic-PD. **(D)** ENCODE data from UCSC Genome Browser showing evidence of regulatory sequences.

active enhancers of gene expression [44]. We did not find compelling evidence for the PD-associated SNPs being eQTL. The expression data and *in-silico* methods for combining disease-association and expression data are evolving rapidly. Studies agree that eQTL patterns vary across tissues and by cell type [30,45,46]. Thus future studies will require not only more enriched eQTL datasets (only a fraction of regulatory elements has been identified) but also a larger variety of tissues and specific cell types.

Prior GWAS have identified over 20 susceptibility loci for PD (Additional file 3), of which *SNCA*, *MAPT* and

HLA have the strongest signals in our data. There is compelling evidence to suggest that association of PD with *SNCA* and *HLA* involves variations in gene expression [47-49]. In contrast, there is little information on the underlying mechanism of the association of PD with *MAPT*. Here, we detected distinct patterns in Sporadic and Familial subtypes. *SNCA* was associated with All-PD, *MAPT* was associated primarily with Familial-PD, and *HLA* was associated only with Sporadic-PD. We also explored differential associations with Familial vs. Sporadic-PD for all genome-wide significant signals from published studies

(Additional file 3). These distinct patterns can help to generate new hypotheses and to select subtypes of PD for specific research questions. The result for *SNCA* would suggest an important and central role for this gene in all forms of PD, which is not a novel thought, considering that α -synuclein is a major component of Lewy-bodies and a ubiquitous diagnostic hallmark of PD. We hypothesize that the stronger association of *MAPT* with Familial-PD is indicative of an incompletely penetrant genetic factor, and that pedigrees may be more informative in studies of *MAPT* and PD. Whereas the seemingly exclusive association of *HLA* with Sporadic-PD suggests that that risk allele is not a cause of disease, rather it might be a genetic susceptibility to an environmental agent. Thus, we suggest that consideration of PD-relevant *HLA* markers may be particularly relevant to studies of epidemiology, exposures and infectious origins of PD.

Conclusions

We discovered and replicated a new locus for Sporadic-PD which had escaped detection in un-stratified GWAS. This demonstrates that by stratifying on a key variable the power gained due to diminished heterogeneity can sometimes outweigh the power lost to reduced sample size. We also detected distinct patterns of disease associations for previously established PD susceptibility genes, which gives an insight to the genetic architecture of the disease and could aid in the selection of appropriate study population for future studies.

Availability of supporting data

The data sets supporting the results for the discovery phase (NGRC dataset) were generated by our group and are publicly available in the NCBI dbGaP repository, phs000196.v2.p1, (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000196.v2.p1). The data sets supporting the results for the replication phase (NINDS dataset) were in part generated by us by direct genotyping of DNA samples from PD cases which we purchased from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>) and in part generated by Simon-Sanchez et al. [2] and are publicly available in the NCBI dbGaP repository, phs000089.v3.p2 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000089.v3.p2).

Additional files

Additional file 1: Table S1. NGRC Subject characteristics.

Additional file 2: Table S2. Association of rs2338971 (chromosome 1p21) with Sporadic-PD was consistent across disease- and study-related strata.

Additional file 3: Table S3. Previously reported genome-wide significant signals stratified by family history.

Abbreviations

PD: Parkinson's disease; GWAS: Genome-wide association study; NGRC: NeuroGenetics Research Consortium; NINDS: National Institutes of Neurologic Disease and Stroke; SNP: Single nucleotide polymorphism; kb: Kilo-base pairs; bp: Base pair; Mb: Mega-base pairs; LD: Linkage disequilibrium; eQTL: Expression quantitative trait locus.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EMH performed data analysis and helped write the manuscript. WTW assisted with data analysis. THH assisted with data analysis. SAF helped recruit subjects for the study. CPZ helped recruit subjects for the study. HP designed the study, coordinated data collection, supervised data analysis and wrote the manuscript. All authors read and approved the final manuscript.

Funding

We would like to acknowledge the persons with PD, their families and healthy volunteers who participated in this study. The project was supported by Award Number R01NS36960 from the National Institute of Neurological Disorders And Stroke. Additional support was provided by a Global Genetic Consortium Grant from the Michael J Fox Foundation for Parkinson's Disease Research, Merit Review Award from the Department of Veterans Affairs (1I01BX000531), National Institutes of Aging (P30AG08017), Office of Research & Development, Clinical Sciences Research & Development Service, Department of Veteran Affairs, The Intramural Research Program of the NIH at National Library of Medicine, and the Close to the Cure Foundation. Genotyping services were provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This study used samples from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>), as well as clinical data. Funding for NINDS-Genome-Wide Genotyping in Parkinson's Disease which generated the GWAS used for replication was provided by NINDS, and the GWAS data were obtained from the NINDS database at (<http://www.ncbi.nlm.nih.gov/gap>) accession number phs000089.v3.p2. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author details

¹Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, NY, USA. ²Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA. ³VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA, USA. ⁴Department of Biomedical Science, School of Public Health, State University of New York, Albany, NY, USA.

Received: 26 April 2013 Accepted: 3 February 2014

Published: 10 February 2014

References

- Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, et al: **Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease.** *Nat Genet* 2009, **41**:1303–1307.
- Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, et al: **Genome-wide association study reveals genetic risk underlying Parkinson's disease.** *Nat Genet* 2009, **41**:1308–1312.
- Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, Kay DM, Doheny KF, Paschall J, Pugh E, et al: **Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease.** *Nat Genet* 2010, **42**:781–785.
- Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, et al: **Web-based genome-wide**

- association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet* 2011, **7**:e1002141.
5. Pankratz N, Beecham GW, DeStefano AL, Dawson TM, Doherty KF, Factor SA, Hamza TH, Hung AY, Hyman BT, Iverson AJ, et al: **Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2.** *Ann Neurol* 2012, **71**:370–384.
6. Hamza TH, Chen H, Hill-Burns EM, Rhodes SL, Montimurro J, Kay DM, Tenesa A, Kusel VI, Sheehan P, Eaaswarkhanth M, et al: **Genome-wide gene-environment study identifies Glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee.** *PLoS Genet* 2011, **7**:e1002237.
7. Hill-Burns EM, Singh N, Ganguly P, Hamza TH, Montimurro J, Kay DM, Yearout D, Sheehan P, Frodey K, McLearn JA, et al: **A genetic basis for the variable effect of smoking/nicotine on Parkinson's disease.** *Pharmacogenomics J* 2013, **13**:530–537.
8. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, et al: **Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.** *Science* 1997, **276**:2045–2047.
9. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, et al: **alpha-Synuclein locus triplication causes Parkinson's disease.** *Science* 2003, **302**:841.
10. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ, Calne DB, et al: **Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.** *Neuron* 2004, **44**:601–607.
11. Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, van der Brug M, de Munain AL, Aparicio S, Gil AM, Khan N, et al: **Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease.** *Neuron* 2004, **44**:595–600.
12. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N: **Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.** *Nature* 1998, **392**:605–608.
13. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z, Del Turco D, Bentivoglio AR, Healy DG, et al: **Hereditary early-onset Parkinson's disease caused by mutations in PINK1.** *Science* 2004, **304**:1158–1160.
14. Bonifati V, Rizzo P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, Squitieri F, Ibanez P, Joosse M, et al: **Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism.** *Science* 2003, **299**:256–259.
15. Ramirez A, Heimbach A, Grundemann J, Stiller B, Hampshire D, Cid LP, Goebel J, Mubaidin AF, Wriekat AL, Roeper J, et al: **Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase.** *Nat Genet* 2006, **38**:1184–1191.
16. Vilarino-Guell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, Soto-Ortolaza AI, Cobb SA, Wilhoite GJ, Bacon JA, et al: **VPS35 mutations in Parkinson disease.** *Am J Hum Genet* 2011, **89**:162–167.
17. McCulloch CC, Kay DM, Factor SA, Samii A, Nutt JG, Higgins DS, Griffith A, Roberts JW, Leis BC, Montimurro JS, et al: **Exploring gene-environment interactions in Parkinson's disease.** *Hum Genet* 2008, **123**:257–265.
18. Ritz BR, Manthripragada AD, Costello S, Lincoln SJ, Farrer MJ, Cockburn M, Bronstein J: **Dopamine transporter genetic variants and pesticides in Parkinson's disease.** *Environ Health Perspect* 2009, **117**:964–969.
19. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, Brice A, Aasly J, Zabetian CP, Goldwurm S, et al: **Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study.** *Lancet Neurol* 2008, **7**:583–590.
20. Lesage S, Anheim M, Condroyer C, Pollak P, Durif F, Dupuits C, Viallet F, Lohmann E, Corvol JC, Honore A, et al: **Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease.** *Hum Mol Genet* 2011, **20**:202–210.
21. Kay DM, Bird TD, Zabetian CP, Factor SA, Samii A, Higgins DS, Nutt J, Roberts JW, Griffith A, Leis BC, et al: **Validity and utility of a LRRK2 G2019S mutation test for the diagnosis of Parkinson's disease.** *Genet Test* 2006, **10**:221–227.
22. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ: **The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service.** *Brain* 2002, **125**:861–870.
23. Howie BN, Donnelly P, Marchini J: **A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.** *PLoS Genet* 2009, **5**:e1000529.
24. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM: **GenABEL: an R library for genome-wide association analysis.** *Bioinformatics* 2007, **23**:1294–1296.
25. Neter J, Kunter M, Nachtsheim C, Wasserman W: *Applied Linear Statistical Models*. 4th edition. New York, NY: McGraw-Hill/Irwin; 1996.
26. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**:263–265.
27. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, Boehnke M, Abecasis GR, Willer CJ: **LocusZoom: regional visualization of genome-wide association scan results.** *Bioinformatics* 2010, **26**:2336–2337.
28. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, Kaleem M, Leung D, Bryden L, Nath P, et al: **A survey of genetic human cortical gene expression.** *Nat Genet* 2007, **39**:1494–1499.
29. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulet CM, Denny TN, Goldstein DB: **Tissue-specific genetic control of splicing: implications for the study of complex traits.** *PLoS Biol* 2008, **6**:e1.
30. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, et al: **Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain.** *PLoS Genet* 2010, **6**:e1000952.
31. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahoul AG, Herman MM, Weinberger DR, Kleinman JE: **Temporal dynamics and genetic control of transcription in the human prefrontal cortex.** *Nature* 2011, **478**:519–523.
32. Liu C, Cheng L, Badner JA, Zhang D, Craig DW, Redman M, Gershon ES: **Whole-genome association mapping of gene expression in the human prefrontal cortex.** *Mol Psychiatry* 2010, **15**:779–784.
33. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, et al: **Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes.** *Genome Res* 2006, **16**:55–65.
34. The ENCODE Project Consortium: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.
35. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, et al: **Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis.** *Nature* 2011, **476**:214–219.
36. Spiegel S, Milstien S: **The outs and the ins of sphingosine-1-phosphate in immunity.** *Nat Rev Immunol* 2011, **11**:403–415.
37. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, et al: **Crystal structure of a lipid G protein-coupled receptor.** *Science* 2012, **335**:851–855.
38. Kokovay E, Wang Y, Kusek G, Wurster R, Lederman P, Lowry N, Shen Q, Temple S: **VCAM1 is essential to maintain the structure of the SVZ niche and acts as an environmental sensor to regulate SVZ lineage progression.** *Cell Stem Cell* 2012, **11**:220–230.
39. Kirschke CP, Huang L: **ZnT7, a novel mammalian zinc transporter, accumulates zinc in the Golgi apparatus.** *J Biol Chem* 2003, **278**:4096–4102.
40. Cuajungco MP, Lees GJ: **Zinc metabolism in the brain: relevance to human neurodegenerative disorders.** *Neurobiol Dis* 1997, **4**:137–169.
41. Dexter DT, Jenner P, Schapira AH, Marsden CD: **Alterations in levels of iron, ferritin, and other trace metals in neurodegenerative diseases affecting the basal ganglia. The Royal Kings and Queens Parkinson's Disease Research Group.** *Ann Neurol* 1992, **32** Suppl:S94–100.
42. Tarohda T, Ishida Y, Kawai K, Yamamoto M, Amano R: **Regional distributions of manganese, iron, copper, and zinc in the brains of 6-hydroxydopamine-induced parkinsonian rats.** *Anal Bioanal Chem* 2005, **383**:224–234.
43. Noh J, Chang SY, Wang SY, Chung JM: **Dual function of Zn²⁺ on the intrinsic excitability of dopaminergic neurons in rat substantia nigra.** *Neuroscience* 2011, **175**:85–92.
44. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci USA* 2010, **107**:21931–21936.
45. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, et al: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325**:1246–1250.

46. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, *et al*: **Mapping cis- and trans-regulatory effects across multiple tissues in twins.** *Nat Genet* 2012, **44**:1084–1089.
47. Devine MJ, Gwinn K, Singleton A, Hardy J: **Parkinson's disease and alpha-synuclein expression.** *Mov Disord* 2011, **26**:2160–2168.
48. Mata IF, Shi M, Agarwal P, Chung KA, Edwards KL, Factor SA, Galasko DR, Ginchina C, Griffith A, Higgins DS, *et al*: **SNCA variant associated with Parkinson disease and plasma alpha-synuclein level.** *Arch Neurol* 2010, **67**:1350–1356.
49. Wissemann WT, Hill-Burns EM, Zabetian CP, Factor SA, Patsopoulos N, Hoglund B, Holcomb C, Donahue RJ, Thomson G, Erlich H, *et al*: **Association of Parkinson disease with structural and regulatory Variants in the HLA region.** *Am J Hum Genet* 2013, **93**:984–993.

doi:10.1186/1471-2164-15-118

Cite this article as: Hill-Burns *et al*: Identification of a novel Parkinson's disease locus via stratified genome-wide association study. *BMC Genomics* 2014 **15**:118.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

